UNIVERSITAS
GADJAH MADA
Model Hybrid MICE-ElasticNet dan Denoising Diffusion Implicit Model untuk Imputasi Data Tabular Numerik
RIKA FEBRIANTI, Prof. Dr. Ing. Mhd. Reza MI Pulungan, S.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Abuzaid, A., & Alkrunz, I. (2024). A Comparative Study On Univariate Outlier Winsorization Methods In Data Science Context. *Statistica Applicata*, *36*(1), 85–99. https://doi.org/10.26398/IJAS.0036-004

Alwateer, M., Atlam, E.-S., El-Raouf, M. M. A., Ghoneim, O. A., & Gad, I. (2024). Missing Data Imputation: A Comprehensive Review. *Journal of Computer and Communications*, *12*(11), 53–75. https://doi.org/10.4236/jcc.2024.1211004

Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In *Journal of Machine Learning Research* (Vol. 13). http://scikit-learn.sourceforge.net.

Cawley, G. C., & Talbot, N. L. C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. In *Journal of Machine Learning Research* (Vol. 11).

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Cheng, S., Osman, N., Qu, S., & Ballan, L. (2024). FastSTI: A Fast Conditional Pseudo Numerical Diffusion Model for Spatio-Temporal Traffic Data Imputation. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2024.3469240

Deng, Y., Chang, C., Ido, M. S., & Long, Q. (2016). Multiple Imputation for General Missing Data Patterns in the Presence of High-dimensional Data. *Scientific Reports*, *6*. https://doi.org/10.1038/srep21689

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00516-9

Fatyanosa, T. N., Firdausanti, N. A., Prayoga, P. H. N., Kuriu, M., Aritsugi, M., & Mendonça, I. (2024). Meta-learning for vessel time series data imputation method recommendation. *Expert Systems with Applications*, *251*. https://doi.org/10.1016/j.eswa.2024.124016

Ferri, P., Romero-Garcia, N., Badenes, R., Lora-Pablos, D., Morales, T. G., Gómez de la Cámara, A., García-Gómez, J. M., & Sáez, C. (2023). Extremely missing numerical data in Electronic Health Records for machine learning can be managed through simple imputation methods considering informative missingness: A comparative of solutions in a COVID-19 mortality case study. *Computer Methods and Programs in Biomedicine*, *242*. https://doi.org/10.1016/j.cmpb.2023.107803

Gretton, A., Borgwardt, K. M., Rasch, M. J., Smola, A., Schölkopf, B., & Smola GRETTON, A. (2012). A Kernel Two-Sample Test Bernhard Schölkopf. In *Journal of Machine Learning Research* (Vol. 13). www.gatsby.ucl.ac.uk/

Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models*. http://arxiv.org/abs/2006.11239

UNIVERSITAS
GADJAH MADA
Model Hybrid MICE-ElasticNet dan Denoising Diffusion Implicit Model untuk Imputasi Data Tabular Numerik
RIKA FEBRIANTI, Prof. Dr. Ing. Mhd. Reza MI Pulungan, S.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*(2), 105–115. https://doi.org/10.1016/j.artmed.2010.05.002

Jiang, D., & Zhang, S. (2025). A filling method for missing soft measurement data based on a conditional denoising diffusion model. *Journal of Computational Science*, *85*. https://doi.org/10.1016/j.jocs.2025.102531

Khan, H., Rasheed, M. T., Liu, H., & Zhang, S. (2024). High-order polynomial interpolation with CNN: A robust approach for missing data imputation. *Computers and Electrical Engineering*, *119*. https://doi.org/10.1016/j.compeleceng.2024.109524

Kotelnikov, A., Baranchuk, D., Rubachev, I., & Babenko, A. (2023). *TabDDPM: Modelling Tabular Data with Diffusion Models*.

Liu, M., Huang, H., Feng, H., Sun, L., Du, B., & Fu, Y. (2023). *PriSTI: A Conditional Diffusion Framework for Spatiotemporal Imputation*. http://arxiv.org/abs/2302.09746

Liu, Y., Ajanthan, T., Husain, H., & Nguyen, V. (2024). *Self-Supervision Improves Diffusion Models for Tabular Data Imputation*. https://doi.org/10.1145/3627673.3679829

Nakagawa, S., & Freckleton, R. P. (2008). *Missing inaction: the dangers of ignoring missing data*. http://www.multiple-imputation.com,

Ngueilbaye, A., Wang, H., Mahamat, D. A., & Junaidu, S. B. (2021). Modulo 9 model-based learning for missing data imputation. *Applied Soft Computing*, *103*. https://doi.org/10.1016/j.asoc.2021.107167

Nichol, A., & Dhariwal, P. (2021). *Improved Denoising Diffusion Probabilistic Models*. http://arxiv.org/abs/2102.09672

Osman, E. G. A. (2025). Entropy based solutions for detecting and treating multicollinearity in econometrics. *Discover Analytics*, *3*(1). https://doi.org/10.1007/s44257-025-00047-0

Rubin, D. B. (1976). *Accessibility support: If you experience accessibility issues with this file, report them here Inference and Missing Data* (Vol. 63, Number 3).

Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purposes: a multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, *88*(15), 2909–2930. https://doi.org/10.1080/00949655.2018.1491577

Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics and Data Analysis*, *47*(1), 165–174. https://doi.org/10.1016/j.csda.2003.10.012

Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Torné, R. V., Sala, E., Lió, P., Patel, M., Preller, J., Lió, P., Walton, N., Holzer, M., Wassin, M., Escudero Sánchez, L., Babar, J., Prosch, H., … Schönlieb, C. B. (2023). The impact of imputation quality on machine learning classifiers for datasets with missing values. *Communications Medicine*, *3*(1). https://doi.org/10.1038/s43856-023-00356-z

UNIVERSITAS
GADJAH MADA
Model Hybrid MICE-ElasticNet dan Denoising Diffusion Implicit Model untuk Imputasi Data Tabular Numerik
RIKA FEBRIANTI, Prof. Dr. Ing. Mhd. Reza MI Pulungan, S.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Song, J., Meng, C., & Ermon, S. (2020). *Denoising Diffusion Implicit Models*. http://arxiv.org/abs/2010.02502

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *Journal of Statistical Software mice: Multivariate Imputation by Chained Equations in R* (Vol. 45). http://www.jstatsoft.org/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Wang, X., Zhang, H., Wang, P., Zhang, Y., Wang, B., Zhou, Z., & Wang, Y. (2023). An Observed Value Consistent Diffusion Model for Imputing Missing Values in Multivariate Time Series. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2409–2418. https://doi.org/10.1145/3580305.3599257

Wang, Y., Xu, X., Hu, L., Liu, J., Yan, X., & Ren, W. (2024). Continuous imputation of missing values in time series via Wasserstein generative adversarial imputation networks and variational auto-encoders model. *Physica A: Statistical Mechanics and Its Applications*, *647*. https://doi.org/10.1016/j.physa.2024.129914

Wen, Y., Yi, K., Ke, J., & Shen, Y. (2024). *DiffImpute: Tabular Data Imputation With Denoising Diffusion Probabilistic Model*. http://arxiv.org/abs/2403.13863

Wibisono, A., Denny, Mursanto, P., & See, S. (2024). Natural generative noise diffusion model imputation. *Knowledge-Based Systems*, *301*. https://doi.org/10.1016/j.knosys.2024.112310

Xiao, C., Jiang, X., Du, X., Yang, W., Lu, W., Wang, X., & Chetty, K. (2025). Boundary-enhanced time series data imputation with long-term dependency diffusion models. *Knowledge-Based Systems*, *310*. https://doi.org/10.1016/j.knosys.2024.112917

Xu, H., Liu, Z., Wang, H., Li, C., Niu, Y., Wang, W., & Liu, X. (2024). Denoising Diffusion Straightforward Models for Energy Conversion Monitoring Data Imputation. *IEEE Transactions on Industrial Informatics*. https://doi.org/10.1109/TII.2024.3413349

Zhang, X., & Zhou, P. (2024). A transferred spatio-temporal deep model based on multi-LSTM auto-encoder for air pollution time series missing value imputation. *Future Generation Computer Systems*, *156*, 325–338. https://doi.org/10.1016/j.future.2024.03.015

Zhao, W., Chen, Z., Xie, P., Liu, J., Hou, S., Xu, L., Qiu, Y., Wu, D., Xiao, J., & He, K. (2024). Multi-task oriented diffusion model for mortality prediction in shock patients with incomplete data. *Information Fusion*, *105*. https://doi.org/10.1016/j.inffus.2023.102207

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. In *J. R. Statist. Soc. B* (Vol. 67, Number 2). https://academic.oup.com/jrsssb/article/67/2/301/7109482